JAHR | Vol. 12/2 | No. 24 | 2021

Review article / Pregledni članak
Received / Zaprimljen: 11. 06. 2021.

https://doi.org/10.21860/j.12.2.9

Hyeongjoo Kim *, Jinkyu Jeong **

On a Possibility of Artificial Reason: J. McCarthy, I. Kant, and A. Turing

SUMMARY

The purpose of this study is to explore the possibility of reconciliation between Kant's transcendental idealism and McCarthy's epistemological point of view on artificial intelligence, which are at first glance likely to be considered contradictory. For this, characterizing the standpoint of J. McCarthy, who coined the word 'artificial intelligence' as scientific realism and that of A. Turing, who provided a crucial thought experiment that shaped the contemporary conception of artificial intelligence as behaviorism, we shall compare these two standpoints with the transcendental idealism of I. Kant, who conferred on us a monumental indicator for understanding the human reason. Through this comparison, we shall argue that scientific realism, which is currently a prominent philosophical standpoint of artificial intelligence, is not compatible with Kant's transcendental idealism but assumes a standpoint strikingly analogous to behaviorism. Nevertheless, we shall also argue that once transcendental idealism is looked at from the viewpoint of behaviorism, scientific realism can be seen as compatible with transcendental idealism. This compatibility we name the possibility of artificial reason in this paper.

Keywords: Artificial Intelligence, I. Kant, Transcendental Idealism, Behaviorism, Scientific Realism.

1. Introduction: Birth of ai on background of technology-criticism

"We will, as we say, *get* technology *spiritually in hand*. We will master it. The will to mastery becomes all the more urgent the more technology threatens to slip from human control." (Heidegger, 1977, p. 5)

* Humanities Research Institute, Chung-Ang University. Seoul, Republic of Korea. E-mail: godwithhj@cau.ac.kr. ORCID: https://orcid.org/0000-0001-8049-6277.

** Minerva College, Hankuk University of Foreign Studies, Seoul, Republic of Korea, E-mail: twojkent@gmail.com.

Correspondence Address: Jinkyu Jeong, Minerva College, Hankuk University of Foreign Studies, 107, Imun-ro, Dongdaemun-gu, Seoul, 02450, Republic of Korea. E-mail: twojkent@gmail.com.

This confession by M. Heidegger, who defined the 1950s as the age of nuclear power, is not so unfamiliar to us facing the age of artificial intelligence. In fact, the question of technology after World War II dominated German thought at the time. For example, K. Jaspers (2010, p. 115) said, "Technology is per se neither good nor evil, but it can be used for either good or evil", and Heidegger (1977, p. 4) responds, "Everywhere we remain unfree and chained to technology, whether we passionately affirm or deny it." According to him, technology is no longer just a means. In his view (Heidegger, 1977, p. 4), a person like Jaspers underestimates the power of the spirits, a technique that has already begun breathing on its own, or more precisely, has recently been breathing heavily. "The essence of technology is by no means anything technological." Meanwhile, his teacher, Husserl, was also concerned about shrinking the spiritual world due to the expansion of technology and science. He diagnoses the atmosphere of the academia at the time with a positivist unilateral approach to explaining life world in dry arithmetic language as "The crisis of a science" (Husserl, 1970, p. 42). He criticized the scientific worldview by claiming that positivism, which tries to show everything with numbers or sets, becomes a shield that hides the true nature of life. Also, W. Dilthey's term "Human Sciences (Geisteswissenschaften)" (1989, p. 55) can be understood as the same. He emphasizes the necessity of its universal methodology to secure the unique domain of the humanities from the natural science, which was expanding its power by using mathematics. This strategy is derived from I. Kant, who has been stigmatized as a breakthrough in metaphysics regardless of his own will.

So far, the situation in Germany in 1950 is as follows, and the situation in the United States at the same time. To explain, the concept of "artificial intelligence" was first established by engineers at the Dartmouth Conference in 1956, and this is the period when philosophers began to study the concept in earnest. However, if we think about it in reverse, it can be said that the prevailing academic climate at the time formed the background for the birth and design of "artificial intelligence". In fact, the philosophy that dominated the United States at the time was an empirical positivist philosophy that stemmed from the irritation of speculative philosophy.

Positive philosophy tried to redefine human reason more clearly at the same time as the metaphysical elimination of non-physical entities such as the soul. The spirit (nous) that understanding the logos of nature and the world, the ability to directly confront the divine spirit (Intelligentia), and the working principle of the ego (intellektuelle Anschaung), these are "unspeakable" to them, that is, these are the object of silence. Logical positivism, which emphasized "logic" solely from the numerous functional predicates applied to reason, became the basis for artificial intelligence algorithms by using theorems and proof methods of modern predicate logic. And the philosophy of mind, which understood the problem of object and

cognition as the mind-body problem, became the basis for the idea of establishing an artificial intelligence system for processing stimulation and information. In short, the philosophy of logical positivism was behind the birth of AI.

On a related note, a look at the landscape of American philosophy in approximately 1950 by expanding the horizon makes it clear that pragmatism was also a large mainstream philosophy during the same period. Pragmatists saw "problems" in the events faced by humans and regarded the process of solving them as a living process. Therefore, "problem-solving" itself is the philosophy of pragmatism. This philosophy significantly contributed to defining the role of AI (Kieras & Holyoak, 1987). During the 1920s, when Dewey consolidated his position as the leader of pragmatism, logical positivism appeared on the stage of American philosophy. The exchange between these two camps became increasingly active, and a philosophical common denominator was established. Despite the decisive difference between the two, owing to the pragmatist advocacy of adduction as a third form of reasoning, they have a common denominator as scientifically oriented philosophies. At a more concrete level, they have in common their belief that the true source of knowledge is experience, based on their perception that their predecessors are British empiricists, and in their attitude that philosophy is not a matter of theory but of methodology (Nekrasas, 2001). It is also noteworthy that by the time AI was born during the 1950s, merging the two philosophical trends was underway.

Reflections and criticisms that follow these scientific developments and those technological advancements may typically mean the value judgement after the facts that preceded them. Nonetheless, it may also mean the exposure of the worldview that lurks behind these developments and advancements. These perspectives allow us to appreciate perhaps these facts: these scientific developments and those technological advancements are not necessarily preceded by humanistic reflections. Rather, to the contrary, the worldview is enabled by the humanistic reflections of the time that brought about scientific technologies. On the other hand, it would not be an overstatement to say that these scientific technologies of our time are heavily invested in the developments of Artificial Intelligences (A.I.'s).

In this paper, we attempt to define the worldviews that are generated by these developments and those advancements. For the convenience of discussion, we focus on three notable standpoints. Firstly, the standpoint of J. McCarthy, the computer engineer who first created the term 'artificial intelligence. And I put his worldview in the category of scientific realism; secondly, that of I. Kant, the philosopher who is considered the founding father of German Idealism. The theory is called transcendental idealism. Thirdly, that of A. Turing, the mathematician who first suggested the contemporary conception of artificial intelligence in his classical

essay "Computing Machinery and Intelligence" (1950). I consider this as a kind of behaviorism. We shall finally make clear the relations among these three perspectives. Through this comparison, we shall argue that scientific realism, which is currently a prominent philosophical standpoint of artificial intelligence, is not compatible with Kant's transcendental idealism but assumes a standpoint strikingly analogous to behaviorism. But we shall also argue that once transcendental idealism is looked at through the viewpoint of behaviorism, scientific realism can be seen as compatible with transcendental idealism.

2. McCarthy's scientific realism

As we have seen above, philosophical reflection on artificial intelligence takes place in the category of scientism and technology criticism. Meanwhile, "problem-solving" is never a mission when it comes to defining AI. For example, *Artificial Intelligence: A Modern Approach*, which is the standard of AI textbooks, defines AI as a "problem-solving agent." In other words, AI is a tool that solves problems that require intelligence. However, it is to note, not all positions in science are geared toward solving problems. Semantic realism, which is a type of scientific outlook, supports the coherence theory of truth, whereas semantic anti-realism replaces realistic concepts of truth, such as a guaranteed argument and limit of inquiry, with epistemological concepts. Semantic realism is divided into scientific realism, which believes that all scientific statements have a truth value, and scientific instrumentalism, which regards science only as a tool of scientific inquiry, deferring the allocation of truth values to scientific laws and theories. Scientific realism also includes methodological realism, which regards truth as an important purpose of scientific inquiry, and methodological non-realism, which replaces truth with methodological substitutes, such as a successful prediction, empirical relevance, and problem-solving ability. Therefore "scientific realists in turn include methodological realists who take truth (usually together with information or systematic power) to be an important aim of scientific inquiry and methodological non-realists who replace truth as an aim of science by some methodological surrogate (e.g., successful prediction, empirical adequacy, problem-solving ability)" (Niniluoto, 1986, p. 258).

AI research, whose main focus is on problem-solving rather than a theoretical quest for truth, may have generally evolved on the basis of methodological non-realism and scientific instrumentalism, albeit to varying extents. The quest for truth decreases with increasing importance attached to the drawing of practical results for scientific phenomena. J. McCarthy, the Father of Artificial Intelligence, was a typical researcher who accepted these views as norms.

According to him, artificial intelligence has the means for problem-solving[1], and the ability concerns a particular situation that occurred within the physical external world. The ontological status of artificial intelligence is regarded as having the equivalent status of a human being because the world in which intelligence itself operates is the physical real world that sustains our scientific common-sense. Let us take a look at his direct comment on this.

> "The physical world already containing intelligent machine called people exists. Information about this world is obtainable through the senses and is expressible internally. Our common-sense view of the world is almost right and that is our scientific view. The right way to think about the general problems of metaphysics and epistemology is not to attempt to clear one's own mind of all knowledge and start with 'Cogito ergo sum' and build up from there (McCarthy & Hayes, 1969, p. 6)".

As mentioned above, understanding the world in which the problem to be solved is given to artificial intelligence, that is, the world in which a specific output is required, is based on common sense that there is no doubt about the existence of an external physical object. It presupposes a natural scientific worldview. And he justifies direct knowledge of the physical world, that is, the external world of consciousness, by stating that the knowledge of the physical world acquired through our sense organs can become the intrinsic knowledge of the subject of perception. This can be easily solved and reconstructed as follows.[2]

> The physical world that contains humans is real. Therefore, the physical world does not exist within human cognition but rather exists outside it. However, knowledge about the physical world acquired through the experience of human sense organs becomes human intrinsic knowledge. This is the scientific common sense we generally have, and as a result, it can be said to be correct.

To sum them up: within the world that we inhabit — i.e., the actual world — there is intelligence, and the common sense, for the most part, gets it right about them. One of the most important implications is that space exists outside of our minds. In other words, space actually exists. In this position, it is an overly speculative attitude in opposition to our common-sense to raise doubts about the reality of all sensible

1 In this regard, he says: "We have to say that a machine is intelligent if it solves certain classes of problems requiring intelligence in humans or survives in an intellectually demanding environment" (McCarthy & Hayes, 1969, p. 4).

2 We consider McCarthy's scientific realism, according to which Artificial Intelligence is but one of many intelligences, to be founded on the naïve scientific realism of common sense. The goal to engineer AI's that solves the "problems" — the problem-solving machines — and setting its goal accordingly stems from the scientific realism's worldview. This worldview has been established by J. McCarthy, whose conceptual innovations allowed for the highly sought out notions of Artificial Intelligence. He belonged to the group of theorists with the foundational accomplishments of establishing the notion itself. To appreciate this fact and state it explicitly, we shall consider the view that we call scientific realism. It is also to establish its sophisticated and compounded nature of the topic.

things and further all the things that are thought to occupy space. A statement of existence means nothing more than one of physical existence. Indeed, our common-sense follows this position, and in this regard, we all are scientific realist.

3. Kant's transcendental idealism

We can draw from "2. McCarthy's Scientific Realism", space actually exists outside of our mind. In this position, it is an overly speculative attitude in opposition to our common-sense to raise doubts about the reality of all sensible things and further all the things that are thought to occupy space. A statement of existence means nothing more than one of physical existence. Indeed, scientific common-sense follows this position, and in this regard, we all are an empirical realist who considers that the things in this world exist in reality, even if they can be unobservable. Some of the stuff that exist in the world may not be observable for many different reasons. However, we sometimes reflect on our common-sense. The result of this reflection, upon occasion, may have a logical consistency and bring a significant insight on the subject.

On the other hand, I. Kant is skeptical of the possibility of direct recognition of things that exist outside of his consciousness and limits the realm of certain knowledge to the inside of his consciousness. He refers to this worldview as transcendental idealism, and the worldview he criticizes as the transcendental realism. The transcendental realism is in line with McCarthy's realism, which we are now discussing. To clarify this point, consider the following I. Kant's remarks:

> "I understand by the transcendental idealism of all appearances the doctrine that they are all together to be regarded as mere representations and not as things in themselves, and accordingly that space and time are only sensible forms of our intuition, but not determinations given for themselves or conditions of objects as things in themselves. To this idealism is opposed the transcendental realism, which regards space and time as something given in themselves (independent of our sensibility). The transcendental realist therefore represents outer appearances (if their reality is conceded) as things in themselves, which would exist independently of us and our sensibility and thus would also be outside us according to pure concepts of the understanding" (Kant, 1900ff, A369).

In the passage above, transcendental idealism clearly distinguishes thing-in-itself and phenomenon and limits the area cognized by human intelligence to the phenomenon world merely. According to Kant's own transcendental idealism, however, space and time are nothing but pure forms of our sensible intuition, and thus, strictly speaking, the external world is not real. The world described by transcendental idealism is the world of self-representation. For the transcendental idealist, the only things

that are certain are those representations formed in our consciousness which are phenomenalized things-in-themselves. The cause of this distinction and limitation is the consciousness - immanence of the concept of space-time. According to it, space-time is not an entity that exists outside of human intelligence itself but is a form of human intelligence that enables cognition. Hence, according to Kant's transcendental idealism, we can only "relate to representations" (Kant, 1900ff, A190/B235) that we self-create accepted through space-time.

Meanwhile, I. Kant defines his transcendental realism as a theory that presumes space as given independently of our consciousness and regards things in space as "things-in-themselves". It seems quite evident that the claims of transcendental realism in this definition resemble those of scientific realism considered above. For the "things-in-themselves" in Kant's terms are no other than those physical realities the existence of which scientific realism takes for granted. And the expression "represents outer appearances as things in themselves" is consistent with the basic idea of the scientific realism that our soul and the physical world is ultimately no different, in other world, the real world and the relying world are virtually one. Based on this interpretation, according to Kant's criteria, J. McCarthy, whom we defined above as a scientific realist, can be regarded as the transcendental realist whom I. Kant has criticized.

To clarify this fact, we directly contrast each of the key passages one by one in reference to I. Kant and J. McCarthy, which we have focused on above. I. Kant's sentences are denoted by [K] and J. McCarthy's sentences by [M].

[K] He acknowledges the existence of matter without going out of his consciousness and without assuming more than the certainty of the representation within me, that is, 'I think, therefore I am'.

[M] The right way to think about the general problems of metaphysics and epistemology is not to attempt to clear one's own mind of all knowledge and start with 'Cogito ergo sum' and build up from there.

"Without going out of his consciousness" in [K] has the same meaning as "not to attempt to start with "Cogito ergo sum" and build up from there" in [M]. Based on this, we will simply these two sentences as follow.

[K*] Our knowledge begins and ends with our self-consciousness, "Cogito ergo sum".

[M*] Our knowledge is not an attempt for self-awareness, and therefore does not begin with "Cogito ergo sum".

As is evident in [K*] and [M*], whether or not to acknowledge the self-consciousness of "Cogito ergo sum" is the decisive difference that can distinguish the two. In other

words, it is the crucial difference between empirical realism and transcendental realism. According to I. Kant, it is this self-consciousness that makes man 'intelligence'. In the "deduction" section §25 of the Critique of Pure Reason, which mainly discusses the problem of self-consciousness, he says the following about intelligence:

> "Through which (intuition of manifold in me: HJ. K.) I determine this thought (I think myself: HJ. K.); and I exist as an intelligence that is merely conscious of its faculty for combination". "This spontaneity is the reason I call myself an intelligence."

According to the quote above, first, intelligence is the faculty for combination. Also, as we saw earlier, this holds true for both I. Kant and J. McCarthy. However, I. Kant adds one more essential property of intelligence: the ability to be conscious of this faculty. The spontaneous thinking ability that guarantees the identity of the subject "I think" is also represented with apperception (Apperzeption) as the fundamental ability to unify them collectively at the base of the given perceptions (Perzeptionen) and to make them aware that they are my perceptions. For I. Kant, the human as intelligence is a self-conscious subject, that is, a subject who can constantly be aware and should be aware of what are the knowledge, sensitivities, and judgments given to ones now and in the past. In other words, it is a subject constantly conscious of the fact that the representations and thoughts I am now thinking about certainly belong to me. This is what separates I. Kant from J. McCarthy.

4. Turing's behaviorism

Rather than J. McCarthy mentioned in Chapter 2, A. Turing asked the question of the possibility of artificial intelligence as an intelligent being by first presenting the philosophical thesis of "Can machines think?". The well-known Turing test is a way of asking such questions. It is also well known that the theoretical premise of the answer to that question is behaviorism. (Kim, 2011, p. 158). But to wit, "behaviorism is usually referred to in the singular" (Donohue & Kitchener, 1999, p. 1), but *de facto* should be distinguished from the fact. B. Skinner, whom we have just mentioned, and another behaviorist W. Quine make it plain that the doctrine called behaviorism has been developed in a multitude of manners, not simply as a single doctrine. Broadly speaking, it is divided into psychological behaviorism and philosophical behaviorism. The former is called methodological behaviorism, and the latter is called analytical behaviorism because it is influenced by logical positivism. J. Watson, J. Kantor, K. Spence, E. Tolman, etc. belonged to the first camp, and R. Carnap, C. Hempel, G. Ryle, etc. belonged to the second. These two camps take views different from each other. Oftentimes, the differences were methodological,

but other times, the differences were more substantial. They have focused on different problems over an identical period.

Of course, among the theorists from the same camp they do not agree on every issue either. The specifics and the details gave the distinctive voices from the same camp of theorists. In other words, there does not seem to be a set of sufficient conditions that would allow us to determine behaviorism. All we have are the prototype theory or the family resemblance among these distinct doctrines. On the other hand, these available resources can be used to accommodate and explicate the very notion doctrine of behaviorism in the following manner. That is to say, despite the lack of bountiful resources, the many renditions of behaviorism seem to lead the core notion of behaviorism as follows. We define the essential characteristics of an ism as follows: "Behavior can be described and explained without making ultimate reference to mental events or to internal psychological processes. The sources of behavior are external (in the environment), not internal (in the mind, in the head)" (Graham, 2019).

Based on this, looking at the contents of logical behaviorism in which the theoretical premise of the Turing test is discussed in detail, J. Kim explains the core of logical behaviorism founded by C. Hempel as follows: "Any meaningful psychological statement, that is, a statement purportedly describing a mental phenomenon, can be translated, without loss of content, into a cluster of statements solely about behavioral and physical phenomena" (Kim, 2011, p. 68). The ground of this claim is the thought, that "only behavioral and physical phenomena (including physiological occurrences) are publicly observable" (Kim, 2011, p. 69). According to behaviorism, the reason for judging an action moral and/or ethical should only be grounded on the phenomenal moral nature and/or the phenomenal ethical nature of the action itself. That is, the good will of I. Kant and the practical reason through which the moral laws are to be generated, and the morality that is inherent to the actor him/herself will not be discussed here in this paper. Instead, this allows us to focus on the distinction between the approaches *par* intention and motive, and the approaches *par* result and consequences. This is the externalism, not the essentialism at all. It is to be noted that behaviorism implicitly assumes a dichotomy. Behaviorism ignores the motive and/or intention behind the action; it instead focuses solely on the *apparent* behavior. This leads us to assume that behaviorism takes what goes on inside the actor's internal mechanism for granted. This seems true, at least from a logical point of view. With this, let us discuss Turing's view.

His argument for this view seems to be rather simple. We address his Sentences.

> According to the most extreme form of this view, the only way by which one could be sure that machine thinks is to be the machine and feel oneself thinking. One could

> then describe these feelings to the world, but of course, no one would be justified in taking any notice. Likewise, according to this view, the only way to know what a man thinks is to be that particular man. It is, in fact, the solipsist point of view. It may be the most logical view to hold, but it makes communication of ideas difficult. *A is liable to believe* "A thinks but B does not", whilst B believes, "B thinks but A does not". Instead of arguing continually over this point, it is usual to have the polite convention that everyone thinks (Turing, 1990, p. 52).

"As above, A. Turing says, the only way of perfectly confirming that a machine can think is that the questioner becomes that machine. Since that is impossible, our judgment on whether it can really think cannot help depending on the observation of that machine's behaviors, that is, its outputs" (Kim & Byun, 2021, p. 94).[3] We have already noted above that one of the important characteristics of behaviorism is that its attentions are given solely to the apparent phenomena, not the internal workings of the agent. For the behaviorist, the effectiveness of problem-solving has to take precedence over other matters. A. Turing himself acknowledged that it would be logical to attempt figuring out the internal structure of one's making a certain judgment. Nevertheless, he also has made it clear to us that such an endeavor would not allow us to communicate the findings, and in the end, it would hinder the problem-solving. The very test that he has devised — i.e., the Turing test — and its significance lies in this very realization. According to A. Turing, "If a machine seems to be thinking, then we should consider the proposition that the machine thinks to be true" (Kim & Byun, 2021, p. 94). Thus, we may draw the following conclusion from this.

Turing's thought - the Judgment, artificial intelligence thinks, only depends on the fact it appears to think and entirely regardless of whether or not artificial intelligence actually thinks - has something in common with the behaviorist fundamental thesis that the only way of figuring out an agent's intent is to observe her actions. And finally, the implicit assumptions of his, and the dichotomy between the phenomenon and the reality by I. Kant share this common ground between them.

5. Compatibility between transcendental idealism and scientific realism through behaviorism

The argumentation process presented by I. Kant to assert his worldview, transcendental idealism, gives many implications for explaining the artificial intelligence worldview. As stated earlier, transcendental idealism is a theory that "we can never reach direct

3 In this regard, it is meaningful to examine Jaegwon Kim's Turing Thesis. "Turing's Thesis. If two systems are input-output equivalent, they have the same psychological status; in particular, one is mental, or intelligent, just case the other is" (Kim, 2011, p. 158).

awareness of thing-in-themselves (Ding an sich), but only experience the phenomena given to our consciousness as true knowledge". To express his thoughts clearly, he sets up transcendental realism as the opposite concept. In other words, if it is the transcendental idealism to explain the transcendent thing, that is, "things outside me (thing-in-themselves) are mere ideas", the theory that the transcendent thing is real is transcendental realism. And the decisive factor that distinguishes these two theories is the answer to positive or negative answer to the question "Does space-time exist in consciousness?" A negative answer means that the world exists as it is, regardless of the perception of the world, and a positive answer means that the world exists only through its relation to the perception of the world. Therefore, the transcendental realist takes for granted that time and space are outside of consciousness. This again leads to the denial of doubt about the existence of the thing-in-themselves. After examining the distinction between transcendental idealism and transcendental realism, we defined McCarthy's epistemological premise for artificial intelligence as scientific realism through "an acceptance without doubt of the reality of the external world" as a medium, and this was subsumed as transcendental realism. In other words, we use Kant's theory to understand McCarthy's theory as a transcendental realist. Moreover, we took this broadly and inferred that he and J. McCarthy had opposing worldviews, at least from I. Kant's point of view.

On the other hand, we argued the common denominator of behaviorism and Kant's transcendental idealism. A universal characteristic of behaviorism is that it judges that there is truth and value in phenomena, presupposing that it gives up an accurate understanding of the contents of the situation. Through the Turing test, A. Turing argued that when a human and a computer are talking, if the human cannot distinguish whether the computer's speech is from a human or a computer, this is evidence that a machine has intelligence. This argument means that the intelligence of all intelligent beings, including artificial intelligence, is not in essence beyond the phenomenon, but in the phenomenon itself.

What we should note about Turing's behaviorism is that it follows that "phenomena are publicly observable". From the point of view of transcendental idealism, "publicly" expressed from the point of view of behaviorism means universality, which is the criterion for justification of a statement. Therefore, this sentence can be understood to mean the statement that "it is true that a phenomenon is observable". In addition, if we consider the fact that A. Turing negatively evaluates the Cartesian solipsism, as discussed in the previous chapter, this can again be embodied as a statement that "only the recognition of phenomena is justified". This discussion provided us with an opportunity to connect Turing's behaviorism with Kant's transcendental idealism. Comparing the above-discussed sentence on Turing's behaviorism, [H], with the sentence explaining Kant's transcendental idealism, [T],

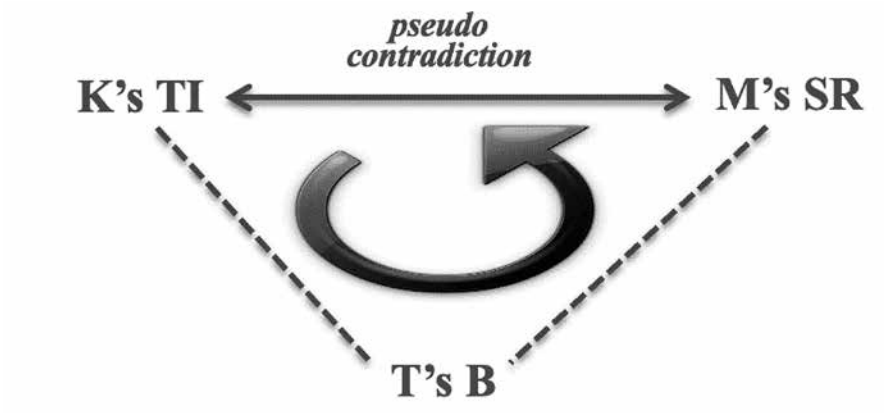
[H] only the recognition of phenomena is justified.

[T] All appearances are to be regarded as mere representations and not as things-in-themselves.

While both behaviorism and transcendental idealism value the recognition of the phenomenon, they take an agnostic or skeptical position about beings beyond the phenomenon.

We take a look at McCarthy's artificial intelligence worldview from a behaviorist point of view, which seems to be a target of criticism from the standpoint of Kant's transcendental idealism. We defined McCarthy's worldview as transcendental realism from the Kantian standpoint and again interpreted it as scientific realism. As stated above, behaviorism is a theory that advocates the value of artificial intelligence represented by the Turing test. The value of artificial intelligence engineering is for pragmatic problem solving, unlike pure science that explores the principles of the natural world. Therefore, A. Turing regarded phenomenal communication as the only criterion for determining the possibility of thinking in a machine. An artificial intelligence device, which can also be expressed as a problem-solving machine, is its entire field of activity in the environment in which it exists. As J. McCarthy mentioned, the physical world already containing intelligent machines is a world where an input value of a problem is given, and a solution called an output value is sent out, that is, the phenomenal world (McCarthy & Hayes, 1969, p. 6). As such, J. McCarthy's scientific realism – precisely the scientific realism we gave to J. McCarthy's theory - and Turing's behaviorism take a common position in that they set problem-solving in the physical phenomenal world as the priority goal that artificial intelligence should perform.

As such, J. McCarthy, I. Kant, and A. Turing's theories all claim to have epistemic value in our world – "in now and here" - rather than metaphysical assumptions such as "substance". We derive this fact through the comparison of transcendental idealism and behaviorism and the comparison of behaviorism and scientific realism. As a result, behaviorism reconciles scientific realism with transcendental idealism that appear to be opposing pairs at first glance. In other words, A. Turing is a conciliator between J. McCarthy and I. Kant. The relationship between these three standpoints can be schematized as follows.



<Picture 1>[4]

6. Conclusion: possibility of artificial reason

We start this paper by examining the philosophical background of the birth of "artificial intelligence" in the 1950s. In the process, we identify J. McCarthy, the creator of the concept of artificial intelligence, I. Kant, a representative anthropocentric, and A. Turing, the developer of the still famous A. Turing test as a representative worldview that can meaningfully view the present age of artificial intelligence. Furthermore, we define each as scientific realism, transcendental idealism, and behaviorism. After that, we argued that scientific realism, a prominent philosophical standpoint of artificial intelligence, is not incompatible with transcendental idealism. Rather, it assumes a standpoint strikingly analogous to behaviorism. It is also argued that scientific realism might be seen as compatible with transcendental idealism if we look at transcendental idealism from the standpoint of behaviorism. What has been argued so far renders plausibility to the following claim. From this idea, we may put on supposition the possibility of harmonization of *rationalism* and the epistemological perspective of *artificial intelligence,* which are likely to be considered contradictory at the first glance. This possibility we name *the possibility of artificial reason* in this paper.

Acknowledgments: This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2021S1A5A8065480).

4 The terms in "Picture 1" is as follows: K's TI is I. Kant's Transcendental Idealism, T's B is A. Turning's Behaviorism, and M's SR is J. McCarthy's Scientific Realism.

References

Allison, H. (2004). *Kant' Transcendental Idealism.* New Haven/London: Yale University Press.

Cassirer, E. (1931). Kant und das Problem der Metaphysik. *Kant-Studien*, *36*(1), 1-26. https://doi.org/10.1515/kant.1931.36.1-2.1

Chakravartty, A. (2017, June 12). Scientific Realism. In E. N. Zalta *The Stanford Encyclopedia of Philosophy,* https://plato.stanford.edu/entries/scientific-realism/

Dilthey, W. (1989). *Einleitung in die Geisteswissenschaften*, In R.A. Makkreel & R. Frithjof (Eds.), *Introduction to Human Sciences*. Princeton: Princeton University Press.

Donohue, W. & Kitchener, R. (Eds.). (1999). *Handbook of Behaviorism*. London: Academic Press.

Emundts, D. (2006, October 21). Idealismus, Transzendentalter. In R. Eisler *Kant Lexikon (Online)*, https://www.textlog.de/

Graham, G. (2019, March 19). Public Health Ethics. In E. N. Zalta *The Stanford Encyclopedia of Philosophy,* https://plato.stanford.edu/entries/behaviorism/#1

Heidegger, M. (1977). Die Frage nach der Technik. In F. W. v. Herrmann (Eds.), *Vorträge und Aufsätze* (pp. 5-36). In W. Lovitt. (Trans.), *The question concerning technology, and other essays*. New York: Garland Pub.

Husserl, E. (1970). *Die Krisis der europäischen Wissenschaften und die transzendentale Phänomenologie: Eine Einleitung in die phänomenologische Philosophie*. In C. David (Trans.). *The Crisis of European Sciences and Transcendental Phenomenology: An Introduction to Phenomenological Philosophy.* Evanston: Northwestern University press.

Jaspers, K. (2010). *Vom Ursprung und Ziel der Geschichte*. In M. Bullock (Trans.). *The Origin and Goal of History*. London: Routledge.

Kant, I. (1900ff.). *Kritik der reinen Vernunft in: gesammelte Schriften (Sog. Akademie-Ausgabe).* Berlin: Walter de Gruyter.

Kim, H. & Byun, S. (2021). Designing and Applying a Moral Turing Test. *Advances in Science, Technology and Engineering Systems Journal*, *6*(2), 93-98. https://doi.org/10.25046/astesj

Kim, J. (2011). *Philosophy of Mind*. Boulder: Westview Press.

Kieras, D. & Holyoak, K. (1987). *Encyclopedia of Artificial Intelligence Vol.1* (Ed. S. Shapiro). Hoboken: John Wileys & Sons.

McCarthy, J. & Hayes, P. (1969), *Some Philosophical Problems from the Standpoint of Artificial Intelligence*, http://jmc.stanford.edu/articles/mcchay69.html

McCarthy, J. (2006, April 25). *What has AI in common with Philosophy?* http://jmc.stanford.edu/articles/aiphil.html

McCarthy, J. (2007, November 12). *What is Artificial Intelligence?* http://jmc.stanford.edu/articles/whatisai.html

Nekrasas, E. (2001). Pragmatism and Positivism. *Problemos, 59*(2), 41-52.

Niiniluoto, I. (1986). Theories, approximations, and idealizations. In R.B. Marcus, G.J.W. Dorn & P. Weingartner (Eds*.), Studies in Logic and the Foundations of Mathematics Vol. 114* (pp. 255-289). Elsevier.

Sellars, W. (1974). …this I or he or it (the thing) which thinks…. In W. Sellars (Eds.), *Essays in philosophy and its history* (pp. 62-90). Dordrecht: Springer.

Strawson, P. F. (1966). *The Bounds of Sense*. London: Methuen & Co. Ltd.

Turing, A. (1990). Computing Machinery and Intelligence. In M.A. Boden (Eds.), *The Philosophy of Artificial Intelligence* (pp. 40-66). Oxford: Oxford University Press.

O mogućnosti umjetnog uma: J. McCarthy, I. Kant i A. Turing

SAŽETAK

Svrha ove studije je istražiti mogućnost pomirenja između Kantova transcendentalnog idealizma i McCarthyjeva epistemološkog stajališta o umjetnoj inteligenciji, koji se na prvi pogled smatraju kontradiktornima. Zbog toga ćemo, karakterizirajući stajalište J. McCarthyja, koji je skovao riječ 'umjetna inteligencija' kao znanstveni realizam, i stav A. Turinga, koji je pružio ključni misaoni eksperiment koji je oblikovao suvremenu koncepciju umjetne inteligencije kao biheviorizma, usporediti ova dva stajališta s transcendentalnim idealizmom I. Kanta, koji nam je pružio monumentalni indikator za razumijevanje ljudskog uma. Pomoću ove usporedbe ustvrdit ćemo da znanstveni realizam, koji je trenutno istaknuto filozofsko stajalište umjetne inteligencije, nije kompatibilan s Kantovim transcendentalnim idealizmom, već zauzima stajalište izrazito analogno biheviorizmu. Ipak, također ćemo tvrditi da se, kada se transcendentalni idealizam promatra kroz stajalište biheviorizma, znanstveni realizam može smatrati kompatibilnim s transcendentalnim idealizmom. Tu kompatibilnost u ovom radu nazivamo mogućnost umjetnog uma.

Ključne riječi: umjetna inteligencija, I. Kant, transcendentalni idealizam, biheviorizam, znanstveni realizam.